



Bhavsar, KA, Singla, J, Al-Otaibi, YD, Song, OY, Zikria, YB and Bashir, AK (2021) Medical diagnosis using machine learning: a statistical review. Computers, Materials and Continua, 67 (1). pp. 107-125. ISSN 1546-2218

Downloaded from: <https://e-space.mmu.ac.uk/627614/>

Version: Published Version

Publisher: Tech Science Press

DOI: <https://doi.org/10.32604/cmc.2021.014604>

Usage rights: Creative Commons: Attribution 4.0

Please cite the published version

Medical Diagnosis Using Machine Learning: A Statistical Review

Kaustubh Arun Bhavsar¹, Jimmy Singla¹, Yasser D. Al-Otaibi², Oh-Young Song^{3,*},
Yousaf Bin Zikria⁴ and Ali Kashif Bashir⁵

¹Lovely Professional University, Jalandhar, 144411, India

²Department of Information Systems, Faculty of Computing and Information Technology in Rabigh,
King Abdulaziz University, Jeddah, 21589, Saudi Arabia

³Software Department, Sejong University, Seoul, 05006, Korea

⁴Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, Korea

⁵Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, M15 6BH, UK

*Corresponding Author: Oh-Young Song. Email: oysong@sejong.edu

Received: 02 October 2020; Accepted: 20 October 2020

Abstract: Decision making in case of medical diagnosis is a complicated process. A large number of overlapping structures and cases, and distractions, tiredness, and limitations with the human visual system can lead to inappropriate diagnosis. Machine learning (ML) methods have been employed to assist clinicians in overcoming these limitations and in making informed and correct decisions in disease diagnosis. Many academic papers involving the use of machine learning for disease diagnosis have been increasingly getting published. Hence, to determine the use of ML to improve the diagnosis in varied medical disciplines, a systematic review is conducted in this study. To carry out the review, six different databases are selected. Inclusion and exclusion criteria are employed to limit the research. Further, the eligible articles are classified depending on publication year, authors, type of articles, research objective, inputs and outputs, problem and research gaps, and findings and results. Then the selected articles are analyzed to show the impact of ML methods in improving the disease diagnosis. The findings of this study show the most used ML methods and the most common diseases that are focused on by researchers. It also shows the increase in use of machine learning for disease diagnosis over the years. These results will help in focusing on those areas which are neglected and also to determine various ways in which ML methods could be employed to achieve desirable results.

Keywords: Decision making; disease diagnosis; machine learning; medical disciplines

1 Introduction

Diagnosis is a way to classify medicines that are fundamental to how a medicine performs its part in society. It is central to the medical system. It organizes disease: defining care choices,



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

forecasting results, and offering an informative mechanism [1]. Appropriate and effective treatment usually involves a thorough diagnosis. The accompanying improvements in diagnostic testing and imaging have certainly improved the entire process of diagnosis. But the human method of scientific judgment leading to correct diagnosis remains key to superior quality and healthy medical services even in this era of rapid technical transition [2]. However, the diagnostic error that harms patient does happen frequently. Generally, multiple factors give rise to diagnostic errors, usually including both perceptual and system-related causes. Certain common factors involve misjudging the significance of observations, misinterpretation, errors originating from heuristics usage, and errors in judgment, particularly when diagnostic hypotheses are developed and assessed [3–5]. Since treatment options are becoming efficient and expensive, the health and financial risk of misdiagnosing an easily curable illness is significantly greater. Thus, there is a loss in improved patient care [6].

These diagnostic errors could be minimized using techniques like fuzzy logic [7], or machine learning (ML) and thus could improve healthcare services. The kind of analytics a clinician can get using ML, at the time of patient treatment, can provide them with more knowledge and, thus, better care [8]. ML tackles the concern of how these systems can be designed that develop with experience continuously. It is known as one of the fastest-growing technical disciplines of today, standing at the junction of computing and analytics and at the heart of artificial intelligence (AI) and data science [9]. Till today, the primary winners of the 21st-century boom in the development of big data, ML, and data science are markets that have been able to obtain such data and employ the workers needed to turn their products. The algorithms built in and around these markets provide considerable potential around improving research in medical and clinical care, particularly provided that clinicians are widely using electronic health records (EHR). Diagnosis and outcome estimation are two fields that gain from the use of ML techniques in the healthcare sector [10]. ML can not only handle varying raw data combinations and apply context weighting but also measure the predictive capacity of any possible combination of factors for determining diagnostic and prognostic components [11]. For example, assisting clinicians for ‘second opinion,’ as based on clinical data, ML models can diagnose aphasia speech type [12], urinary tract infection [13], or even predicting breast cancer [14], among others. The capability to process large data sets far beyond limits of human abilities, and then to efficiently process that data into clinical knowledge that enables doctors to prepare and deliver treatment, eventually leading to improved results, lower medical costs, and enhanced patient satisfaction. ML has the capability and is currently behind the creation of guidelines for precision medicines, treatment counsel, and disease diagnosis [8]. Utilization of these capabilities of ML can even be seen in healthcare internet of things (H-IoT); to analyze and process massive amount of healthcare data generated through sensors [15]. Therefore, extensive research in the context of treatment for specific diseases has been conducted for its usefulness. Hence, the main aim of this paper is to analyze the experiments in which ML approaches are used in relation to different medical fields and diseases to determine their pattern and usefulness in the diagnosis of disease, through a systematic analysis. [Tab. 1](#) represents the uniqueness of our paper using comparative analysis with other published review papers in the medical domain. This paper provides in-depth analysis and results of the use of ML in disease diagnosis. This research paper provides detailed analysis, covering all the major medical domains to the best of our knowledge.

Table 1: Comparative analysis with other published review articles in medical domain

Article	Year	Data types considered	Technologies considered	Medical domains considered	Classification and summarization of existing work	In-depth analysis of reviewed articles
		<ul style="list-style-type: none"> • Imaging • Tabular 	<ul style="list-style-type: none"> • Machine learning 	<ul style="list-style-type: none"> • Cardiology • Critical care • Dermatology • Endocrinology • Gastroenterology • Hepatology • Infectious • Nephrology • Neurology • Oncology • Ophthalmology • Pediatrics • Periodontology • Pulmonology • Rheumatology • Urology • Vascular surgery • Virology. 	<ul style="list-style-type: none"> • Publication year • Authors • Type of articles • Research objective • Inputs and outputs • Problem and research gaps • Findings and results 	<ul style="list-style-type: none"> • Frequency of published articles over the past years • Distribution of academic papers by journal and conference type • Distribution of papers by database providers • Distribution of ML methods applied in published articles • Distribution of ML based on clinical aspects
[16]	2019	✓	×	×	×	×
[17]	2020	×	✓	×	×	×
[18]	2017	✓	✓	×	×	×
[19]	2020	✓	✓	×	×	×
This article		✓	✓	✓	✓	✓

1.1 Comparison with Related Survey Articles

Yanase et al. [16] presented a survey from a computer-aided diagnostic (CAD) system perspective in medicine. The article covers the in-depth workflow of CAD systems and their history. The paper also represents applications in the medical domain from a data type perspective, including tabular, imaging, sound, and signal types of data. Caballe et al. [17] detailed the benefits and limitations of using different ML methods in disease diagnosis. The paper covers classification, regression, and clustering techniques. However, it does not include a summarization of literature and an in-depth analysis of reviewed articles. Jiang et al. [18] surveyed research articles in healthcare from AI perspective. In addition to ML, the paper also covers natural language processing techniques applied in healthcare. The paper covers only three medical domains: cancer, neurology, and cardiology. Schaefer [19] presented an overview of the application of ML in rare diseases. It reviews articles in healthcare covering diagnosis, prognosis, and treatment. None of the articles summarizes the existing work. Also, they cover only a few medical domains and do not provide an in-depth analysis of reviewed articles.

1.2 The Process of Applying Machine Learning in Disease Diagnosis

Medical diagnosis is a complex task largely considered as an empirical task but understood poorly as a cognitive task [20]. Thus, as complex as it may be seen, diagnosis using a computer, i.e., using ML in our case, is divided into multiple steps. The first step of disease diagnosis is data acquisition. This data could be in varied forms, including but not limited to medical interview,

clinical, demographic, imaging, speech, patient historical data, or even heart sound [21–23]. The next step involves processing. In this, the data is prepared, i.e., missing values, dimensionality reduction, dealing with noisy data, and so on is made in this step [24,25]. Next, the target variable and the predictors are identified. This data is then fed to one of the models for training. Once the model is trained, it is then used for diagnosis.

1.3 The Benefits of Using Machine Learning for Disease Diagnosis

Limitations posed due to a large number of overlapping structures and cases, and distractions, tiredness, and limitations with the human visual system, provision of ‘second opinion’ can come handy [26]. This has encouraged the use of CAD systems for diagnostic processes. CAD is a concept that gives equal roles to physicians as well as to the computers, i.e., it assists the physicians in taking the best clinical decisions/practices [27]. Moreover, due to increasing complexity among patients, high diagnostic errors, and availability of a large amount of data, EHR systems are being used to assist in making the clinical decision [28].

With the availability of intelligent tools for data analysis, ML methods help in demystifying interesting relationships in the data [29]. As a second opinion, it could corroborate with clinicians’ decisions or refute it [30]. Integration of ML based tools that monitor continuously increasing volume of data streams for patterns, assisting in decision making for clinicians, or automatically adjusting settings of bedside devices have improved outcomes of patient treatment and substantially reduced the overall cost of treatment [31,32]. On the downside, ML promises to provide the best clinical assistance but so far has not proven useful, according to the article [33,34], probably due to opacity in ML algorithms and analytics. Moreover, data quality and generalizability of the ML models remain amongst the other problems [35,36].

1.4 Article Structure

The paper is organized as follows. Section 2 proposes the methodology employed to carry out this study. It discusses the database chosen and eligibility criteria for the selection of papers. In Section 3, we present the analysis and synthesis of the eligible papers. A discussion of the analysis done is discussed in Section 4. Finally, we draw a conclusion. Fig. 1 presents the taxonomy of this article. Abbreviations and their corresponding full forms used in this article are presented in Tab. 2.

2 Research Methodology

Methodology, in which the author finds relevant studies, selects and investigates those studies, analyzes the data, and summarizes the findings to reach precise conclusions, is called systematic review [37,38]. The use of evidence from dependable research to make healthcare decisions facilitates the use of best practices with lesser mistakes for clinical decision making. Hence, systematic reviews, as well as clinical practice, are considered as the finest source of evidence [39]. The following section includes literature search, study selection, and eligible papers, and extraction and analyzation of data.

2.1 Literature Search

To select relevant and eligible papers for systematic review, six databases were selected in this step.

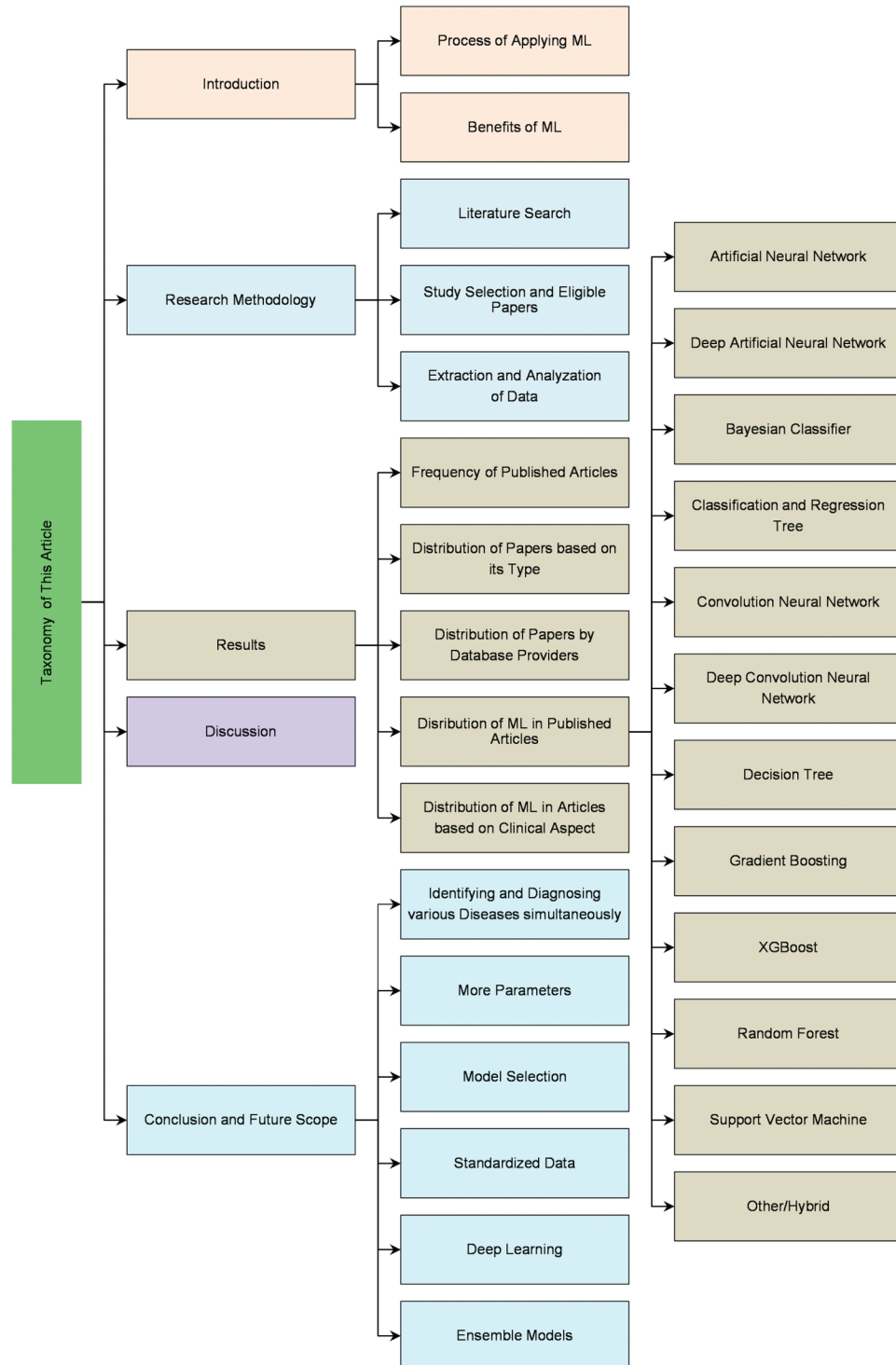


Figure 1: The taxonomy of this article

Table 2: List of abbreviations

Abbr.	Full forms
AI	Artificial intelligence
ANN	Artificial neural network
AUC	Area under curve
BC	Bayesian classifier
CAD	Computer-aided diagnosis
CART	Classification and regression trees
CNN	Convolution neural network
DL	Deep learning
DT	Decision tree
EHR	Electronic health record
FNN	Feedforward neural network
GA	Genetic algorithm
GB	Gradient boosting
LDA	Latent dirichlet allocation
LR	Logistic regression
LSTM	Long short-term memory
ML	Machine learning
NB	Naïve Bayes
PCA	Principal component analysis
PSA	Particle swarm algorithm
RF	Random forest
RNN	Recurrent neural networks
SVM	Support vector machine
XGBoost	Extreme gradient boosting

These databases were: IEEE, PubMed, Science Direct, SciPub, Springer Link, and Web of Science. The articles searched were from the year 2015 up to now. Phrases and keywords such as “disease diagnosis,” “disease diagnosis using machine learning,” “Chronic kidney diagnosis using machine learning,” “Parkinson diagnosis using machine learning,” etc. were used to find relevant articles. The articles were filtered based on relevancy and publication date. From our eligible papers selected, the frequency and number of articles published by publishers are shown in [Tab. 3](#). Accordingly, with 22.73% Elsevier had the highest number of publications. BMC, Hindawi, IEEE, Public Library of Science, and Springer stood second with 6.82% of publications. Nature was ranked third with 4.55% of publications. In comparison, the rest of the publishers ranked fourth with 2.27% of publications each.

2.2 Study Selection and Eligible Papers

Inclusion and exclusion criteria were used to select appropriate and relevant articles. The research only concentrates on disease diagnosis using ML. It excludes paper using fuzzy logic or image processing. Accordingly, articles were screened for selection based on their title and abstract. Only journal and conference papers were considered. Books, book chapters, thesis, reports, review articles, and letters to editors were thus excluded from our research.

Table 3: The frequency of published articles by publisher

Publisher	Articles	Percentage
Avicena publishing	1	2.27
BMC	3	6.82
BMJ	1	2.27
Edusoft publishing	1	2.27
Elsevier	10	22.73
Frontiers Media SA	1	2.27
Hindawi	3	6.82
IEEE	3	6.82
Isfahan cardiovascular research center	1	2.27
Lippincott Williams & Wilkins	1	2.27
MDPI	1	2.27
Mosby Inc.	1	2.27
Nature	2	4.55
Public library of science	3	6.82
Science publications	1	2.27
Springer	3	6.82
	36	81.82

Language, time and article qualities were considered for eligible papers. Thus we selected papers written only in the English language and published from the year 2015 up to now. Our research was focused on including all kinds of medical disciplines. However, diseases related to animals and plants were excluded from it. According to our inclusion criteria, articles using methods and techniques that improved the accuracy of disease diagnosis were included.

2.3 Extraction and Analyzation of Data

The included articles were examined to extract and analyze the data with respect to our research objectives. Thus, to meet our objectives, we analyzed the articles according to frequency of articles over the past years, type of academic papers, according to database providers, and according to ML model employed in those articles.

3 Results

The following section represents the findings and results of the analysis and synthesis of the included articles. This result, which is the outcome of a systematic study of the papers, shows the efficiency of applying ML in disease diagnosis. In the following section, the impact of ML and its use in different medical disciplines is studied.

3.1 The Frequency of Published Articles over the Past Years

Our research includes 44 academic papers that met our inclusion criteria. These 44 papers include research papers as well as conference papers. The frequency of published articles is shown in Fig. 2. The articles included are taken from the year 2015 up till now. The graph indicates that since 2015 there has been a significant rise in published articles. This shows that the research for disease diagnosis using ML has been increasing. In fact, from the included articles, almost

40% were published in the year 2019. Hence, it is evident that researchers are showing interest in applying ML techniques in disease diagnosis.

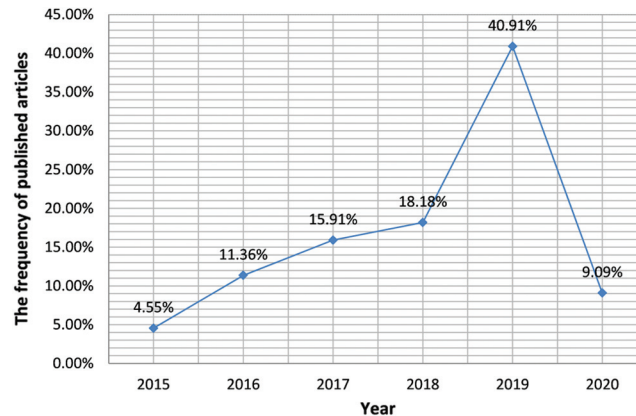


Figure 2: The distribution of papers by publication year

3.2 Distribution of Academic Papers by Journal and Conference Type

Total articles included in this systematic review, including from journals and conferences, are 35 (i.e., 27 from journals and 8 from conference papers). [Fig. 3](#) represents the distribution of papers by publication year and type. As seen in this figure, journal articles published are comparatively higher than conference papers. As we find no conference paper in the year 2015 and 2016 in our chart, we can say that overall fewer articles must have been published in conferences than in journals. However, during 2017 there is a considerable increase in articles published in conferences.

Eligible articles have been categorized by journals and conferences. The distribution of papers by journals is represented in [Tab. 4](#) and by conferences in [Tab. 5](#). From our reviewed articles, almost 81.82% of articles are of journals, and 18.18% of articles are from conferences. ‘Computer Methods and Programs in Biomedicine,’ ‘Computers in Biology and Medicine,’ ‘IEEE Access’ and ‘PLoS ONE’ journals published 6 articles each, which were the highest and concentrated of around 6.82% each. On average, 2.27% of articles were published by each journal.

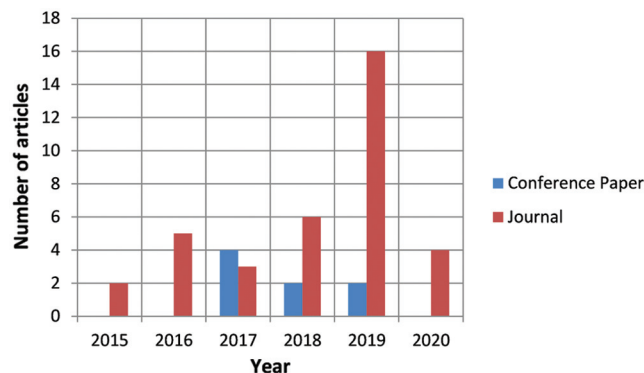


Figure 3: The distribution of papers by publication year and type

Table 4: Distribution of papers based on journal name and publisher

Journal	Count	Percentage	Publisher
Acta informatica medica	1	2.27	Avicena publishing
ARYA atherosclerosis	1	2.27	Isfahan cardiovascular research center
BMC medical informatics and decision making	1	2.27	BMC
BMC medical research methodology	1	2.27	BMC
BRAIN	1	2.27	Edusoft publishing
Brain informatics	1	2.27	Springer
Circulation	1	2.27	Lippincott Williams & Wilkins
Computational and mathematical methods in medicine	1	2.27	Hindawi
Computer methods and programs in biomedicine	3	6.82	Elsevier
Computers in biology and medicine	3	6.82	Elsevier
Critical care	1	2.27	BMC
EBioMedicine	1	2.27	Elsevier
Frontiers in genetics	1	2.27	Frontiers media SA
Gastrointestinal endoscopy	1	2.27	Mosby Inc.
IEEE access	3	6.82	IEEE
Journal of clinical pathology	1	2.27	BMJ
Journal of computer science	1	2.27	Science publications
Journal of healthcare engineering	2	4.55	Hindawi
Multimedia tools and applications	1	2.27	Springer
Nature machine intelligence	1	2.27	Nature
New horizons in translational medicine	1	2.27	Elsevier
Ophthalmology	1	2.27	Elsevier
PLoS ONE	3	6.82	Public library of science
Scientific reports	1	2.27	Nature
Sensors	1	2.27	MDPI
Soft computing	1	2.27	Springer
The American journal of cardiology	1	2.27	Elsevier
Total	36	81.82	

Table 5: Distribution of papers based on conference name and publisher

Conference paper	Count	Percentage	Database provider
2017 13th international conference on signal-image technology & internet-based systems (SITIS)	1	2.27	IEEE
2017 IEEE 19th international conference on e-health networking, applications and services (Healthcom)	1	2.27	IEEE
2017 international conference on intelligent computing and control systems (ICICCS)	1	2.27	IEEE
2017 second international conference on informatics and computing (ICIC)	1	2.27	IEEE
2018 10th computer science and electronic engineering conference (CEEC)	1	2.27	IEEE
2018 IEEE congress on evolutionary computation (CEC)	1	2.27	IEEE
2019 7th mediterranean congress of telecommunications (CMT)	1	2.27	IEEE
2019 IEEE international autumn meeting on power, electronics and computing (ROPEC)	1	2.27	IEEE
Total	8	18.18	

3.3 Distribution of Papers by Database Providers

Searching and selection of papers were made through 6 different databases. These databases and their contribution can be seen in [Tab. 6](#). With 40.91% PubMed was ranked first. It concentrated on 18 papers. Moreover, IEEE was ranked second with 25.00%. Science Direct was ranked third with 22.73%. Springer Link and Web of Science were ranked fourth with 4.55% each. And SciPub was ranked fifth with 2.27%.

3.4 The Distribution of Machine Learning Methods Applied in Published Articles

The objective of this study is to carry out a systematic review of the use of ML methods in disease diagnosis. Also, the distribution of various ML methods for diagnosis is analyzed. From these eligible articles, we can observe that some of these articles incorporated ML to improve the disease diagnosis. Hence, we categorized selected papers into 12 different ML methods, as can be seen in [Tab. 7](#). Among the researchers, support vector machine (SVM) has been ranked one with 22.73%. This shows the efficiency of the SVM method to improve the diagnostic process of diseases. Convolution neural network (CNN) method was ranked second with 15.91%. Other methods, which included proprietary algorithms using ML methods and a combination of various ML methods, ranked third with 13.64%. With 11.36%, random forest (RF) was ranked fourth. Artificial neural network (ANN), deep ANN, and eXtreme gradient boosting (XGBoost) were

ranked fifth with 6.82%. The classification and regression trees (CART) method was ranked sixth with 4.55%. However, bayesian classifier (BC), decision tree (DT), and gradient boosting (GB) were ranked last at 2.27%. This shows that these three ML methods are least preferred in improving the disease diagnosis process.

Table 6: The frequency of published articles by database providers

Database provider	Count of papers	Percentage
IEEE	11	25.00
Pubmed	18	40.91
Science Direct	10	22.73
SciPub	1	2.27
SpringerLink	2	4.55
Web of Science	2	4.55
Total	44	100.00

Table 7: The frequency of applied machine learning methods related to disease diagnosis

Machine learning method	Frequency	Percent	References
Artificial neural network (ANN)	3	6.82	[13,40,41]
Deep artificial neural network (deep ANN)	3	6.82	[42–44]
Bayesian classifier (BC)	1	2.27	[45]
Classification and regression tree (CART)	2	4.55	[46,47]
Convolution neural network (CNN)	7	15.9	[48–54]
Deep convolution neural network (deep CNN)	2	4.55	[55,56]
Decision tree (DT)	1	2.27	[57]
Gradient boosting (GB)	1	2.27	[58]
XGBoost	3	6.82	[59–61]
Random forest (RF)	5	11.36	[62–66]
Support vector machine (SVM)	10	22.73	[67–76]
Other/hybrid	6	13.64	[12,14,77–80]
Total	44	100.00	

We have summarized the distribution of ML methods by year in Fig. 4. Accordingly, we observe that the use of hybrid methods has been increasing over the years (1 in 2015, 2 in 2017, and 4 in 2019) to improve the accuracy of the ML models. From the year 2016, we observe that SVM has always been used over the years, with the highest of 3 articles each in the year 2017 and 2019. This shows its popularity over the years among the researchers. Furthermore, with one article each 2018, 2019, and 2020 XGBoost has shown its consistent use to improve the diagnosis. Also, we observe an increase in the use of the CNN method from 2017.

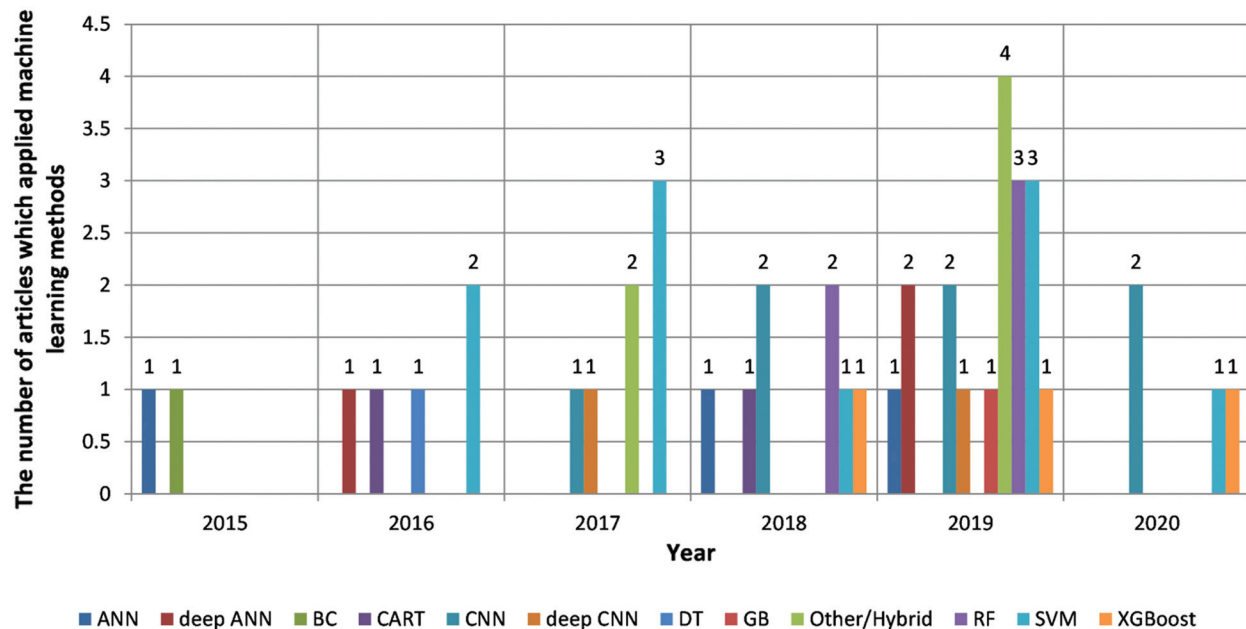


Figure 4: The distribution of machine learning methods by year

3.5 Distribution of ML Methods Applied in Published Articles Based on Clinical Aspects

In the context of disease diagnosis using ML, we would like to know which diseases were considered more. Moreover, in which medical disciplines were researchers more interested is one of the objectives of this research. For this reason, the eligible articles in this research were classified by diseases and the implementation of ML methods. To better understand the distribution of ML for disease diagnosis, we analyzed the articles based on medical disciplines. Fig. 5 represents the pie chart for the frequency of medical disciplines. Based on the diseases, 18 medical disciplines were identified. From Fig. 5, it is observed that 13.64% of studies were carried out in cardiology and endocrinology. Probably due to the large number of diseases associated with them. Infectious disease, oncology, and pulmonology were ranked second with a 9.09%. With a 6.82%, dermatology and nephrology were ranked third. Neurology, rheumatology, and urology were ranked fourth with 4.55%. At last were ranked critical care, gastroenterology, hepatology, ophthalmology, pediatrics, periodontology, vascular surgery, and virology with 2.27% each.

4 Discussion

We conducted this study to review the impact of ML in disease diagnosis. As per our knowledge, fewer articles have been published that systematically analyze academic articles using ML for disease diagnosis. Hence, the results and analysis of this study can be considered to assess the impact of ML in the medical domain and its efficiency in improving the disease diagnosis. This study considered the articles from the year 2015 to 2020. We identified 44 articles applying ML methods to improve disease diagnosis over this period. One of the objectives of this study was to determine which ML methods were used most by researchers for diagnosis, as the answer to this question determines the efficiency of the methods. Hence, the articles were classified accordingly. One of the ways in which articles were classified was based on the number of articles published each year. According to this classification, we observed that the number of publications using

ML for disease diagnosis has been rising over the years. We find that 4.55% of articles were published in 2015, whereas in 2019, 40.91% of articles were published. This article was written in mid-2020. Thus, we were able to retrieve a few articles from this year. This increase in the use of ML methods is due to its efficiency in improving the accuracy and sensitivity of models to give correct results.

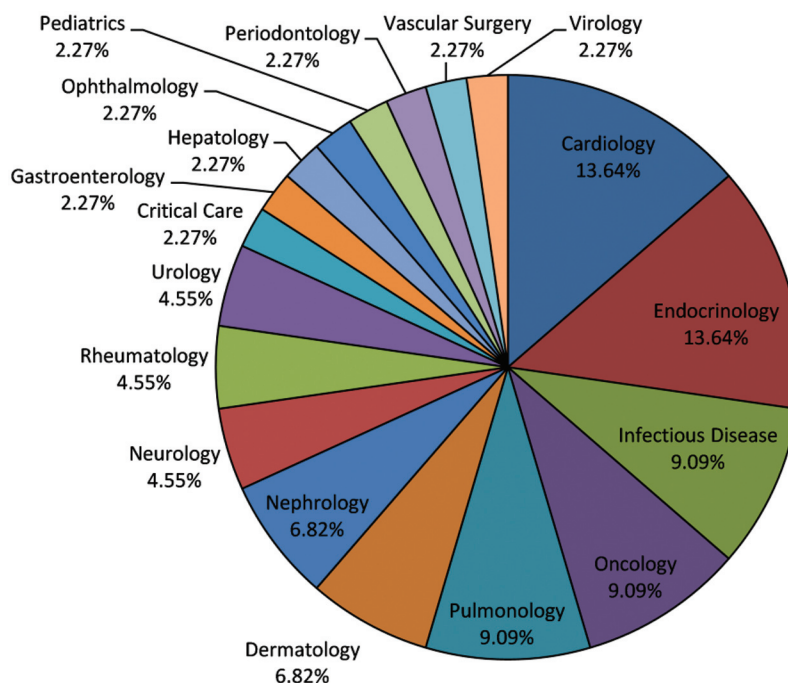


Figure 5: The frequency of medical disciplines

We identified 12 different ML methods that were applied in our eligible papers. Although we say that these 12 methods are mostly used ML methods for disease diagnosis, we limit our findings only to medical diagnosis and do not generalize it. From our analysis, as presented in Fig. 6, we find that researchers prefer SVM, CNN, and RF over other ML methods.

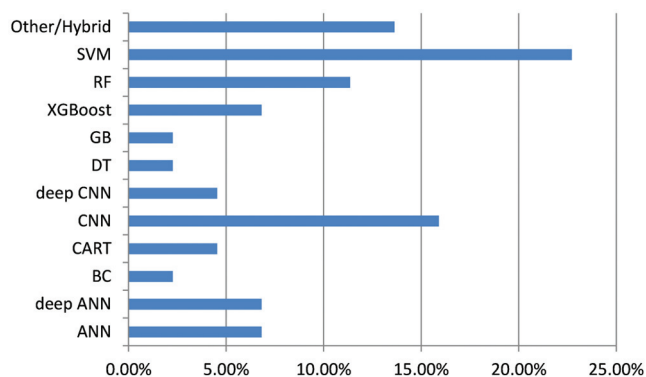


Figure 6: The frequency of machine learning methods impact

However, there is also an increase in the use of hybrid/other methods. This is mainly because using the combination of various methods augments the efficiency of the model. Our study also examined the articles from a medical discipline point of view, i.e., we classified the eligible articles according to medical disciplines. This classification helped us understand which medical disciplines were chosen largely. From this study, it was evident that cardiology and endocrinology had the highest number of publications. This must be due to the fact that most of the diseases come under these two disciplines and also because of the easily available large amount of data to carry out the research. Moreover, going only by diseases explored, we find that variety of ML has been applied to a variety of diseases. This shows the effectiveness of ML in improving the accuracy of disease diagnosis. Thus, we could apply ML in any medical discipline and get the best results.

The findings of this investigation show which diseases and medical disciplines are mostly targeted by researchers and which get neglected. We also find the efficiency of ML methods in disease diagnosis. Therefore, this study could assist researchers in carrying out further work in the medical domain.

5 Conclusion

The main goal of this systematic study was to review the articles using ML for disease diagnosis and, thus, the competence of ML in improving the diagnosis of disease. For the same, we retrieved articles from year 2015 to 2020. We identified six databases including IEEE, PubMed, Science Direct, SciPub, Springer Link, and Web of Science. Further, we classified the articles based on publisher and database. Through this study, we found which databases and publishers are publishing the greatest number of articles relating to ML in disease diagnosis. We also investigated the most used ML methods and their impact on disease diagnosis. Thus, we find that all the studies have shown improvement in their results. We find that using ML not only reduces the overall cost of the treatment and assist clinicians as ‘second opinion,’ but also helps in early detection of diseases having complex structures and patterns. We also identified 12 mostly used ML methods in disease diagnosis and their effectiveness in improving the results. We also investigated the medical disciplines using ML to a large extent. Different ML methods were analyzed to understand their effectiveness in improving disease diagnosis.

Whatsoever, this study has certain limitations. The first limitation is that this systematic review was carried on from year the 2015 to 2020, i.e., for a fixed duration. Also, it has to be noted that this study was carried out up till mid of 2020. But still, through our results, we find that there is growing acceptance and adoption of ML in disease diagnosis over the years. The second limitation of our study is that we did not include articles using fuzzy logic or image processing entirely. In the future, we can include these techniques to get a generalized view and idea of the impact of each of these techniques in disease diagnosis. The third limitation of our study is that our investigation focused solely on the diagnosis of diseases. We did not include articles relating to prognosis or treatment path. In the future, the researchers can investigate the articles to study the impact of ML in prognosis as well as for treatment path.

This study could provide basic knowledge for future studies. We excluded the articles written in other languages and articles other than journals and conference papers. Thus, in the future we can consider neglected resources for investigation as studies of these resources could be valuable. Moreover, we could also identify and diagnose relationship among multiple diseases and diagnose them simultaneously to benefit patients suffering from multiple diseases, investigate with more parameters when building ML models, appropriate selection of models could decrease the time

of implementation, e.g., CNN works better for image data, standardization of data for unbiased results, using deep learning, and ensemble models for better results.

Acknowledgement: Authors wish to thank Dr. Dinesh Grover, Retired Professor, Punjab Agricultural University, Punjab, India for his guidance in writing this review paper.

Funding Statement: This research was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2016-0-00312) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and in part by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW (2015-0-00938) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] A. Jutel, "Sociology of diagnosis: A preliminary review," *Sociology of Health & Illness*, vol. 31, no. 2, pp. 278–299, 2009.
- [2] E. S. Holmboe and S. J. Durning, "Assessing clinical reasoning: Moving from *in vitro* to *in vivo*," *Diagnosis*, vol. 1, no. 1, pp. 111–117, 2014.
- [3] A. Bhasale, "The wrong diagnosis: Identifying causes of potentially adverse events in general practice using incident monitoring," *Family Practice*, vol. 15, no. 4, pp. 308–318, 1998.
- [4] M. L. Graber, N. Franklin and R. Gordon, "Diagnostic error in internal medicine," *Archives of Internal Medicine*, vol. 165, no. 13, pp. 1493–1499, 2005.
- [5] T. K. Gandhi, A. Kachalia, E. J. Thomas, A. L. Puopolo, C. Yoon *et al.*, "Missed and delayed diagnoses in the ambulatory setting: A study of closed malpractice claims," *Annals of Internal Medicine*, vol. 145, no. 7, pp. 488–496, 2006.
- [6] D. Khullar, A. K. Jha and A. B. Jena, "Reducing diagnostic errors – why now?," *The New England Journal of Medicine*, vol. 373, no. 26, pp. 2491–2493, 2015.
- [7] H. Ahmadi, M. Gholamzadeh, L. Shahmoradi, M. Nilashi and P. Rashvand, "Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review," *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 145–172, 2018.
- [8] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future Healthcare Journal*, vol. 6, no. 2, pp. 94–102, 2019.
- [9] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [10] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, "Machine learning in medicine: A practical introduction," *BMC Medical Research Methodology*, vol. 19, no. 1, pp. 1–18, 2019.
- [11] A. M. Darcy, A. K. Louie and L. W. Roberts, "Machine learning and the profession of medicine," *JAMA*, vol. 315, no. 6, pp. 551–552, 2016.
- [12] C. Kohlschein, M. Schmitt, B. Schuller, S. Jeschke and C. J. Werner, "A machine learning based system for the automatic evaluation of aphasia speech," in *2017 IEEE 19th Int. Conf. on e-Health Networking, Applications and Services (Healthcom)*, Dalian, pp. 1–6, 2017.
- [13] I. A. Ozkan, M. Koklu and I. U. Sert, "Diagnosis of urinary tract infection based on artificial intelligence methods," *Computer Methods and Programs in Biomedicine*, vol. 166, pp. 51–59, 2018.
- [14] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani and M. Faisal Nagi, "Automated breast cancer diagnosis based on machine learning algorithms," *Journal of Healthcare Engineering*, vol. 2019, no. 12, pp. 1–11, 2019.

- [15] Y. A. Qadri, A. Nauman, Y. Bin Zikria, A. V. Vasilakos and S. W. Kim, "The future of healthcare internet of things: A survey of emerging technologies," *IEEE Communications, Surveys and Tutorials*, vol. 22, no. 2, pp. 1121–1167, 2020.
- [16] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Systems with Applications*, vol. 138, pp. 112821, 2019.
- [17] N. C. Caballe, J. L. Castillo-Sequera, J. A. Gómez-Pulido, J. M. Gómez-Pulido and M. L. Polo-Luque, "Machine learning applied to diagnosis of human diseases: A systematic review," *Applied Sciences*, vol. 10, no. 15, pp. 1–27, 2020.
- [18] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li *et al.*, "Artificial intelligence in healthcare: Past, present and future," *Stroke and Vascular Neurology*, vol. 2, no. 4, pp. 230–243, 2017.
- [19] J. Schaefer, M. Lehne, J. Schepers, F. Prasser and S. Thun, "The use of machine learning in rare diseases: A scoping review," *Orphanet Journal of Rare Diseases*, vol. 15, no. 1, pp. 145, 2020.
- [20] W. Rogers, B. Ryack and G. Moeller, "Computer-aided medical diagnosis: Literature review," *International Journal of Bio-Medical Computing*, vol. 10, no. 4, pp. 267–289, 1979.
- [21] F. Amato, A. López, E. M. Peña-Méndez, P. Vañhara, A. Hampl *et al.*, "Artificial neural networks in medical diagnosis," *Journal of Applied Biomedicine*, vol. 11, no. 2, pp. 47–58, 2013.
- [22] M. N. Sarkies, K. A. Bowles, E. Skinner, D. Mitchell, R. Haas *et al.*, "Data collection methods in health services research hospital length of stay and discharge destination," *Applied Clinical Informatics*, vol. 6, no. 1, pp. 96–109, 2015.
- [23] J. S. Saczynski, D. D. McManus and R. J. Goldberg, "Commonly used data-collection approaches in clinical research," *The American Journal of Medicine*, vol. 126, no. 11, pp. 946–950, 2013.
- [24] Salvador Gracia, Julian Luengo and Francisco Herrera, *Data Preprocessing in Data Mining*, vol. 72. Switzerland: Springer International Publishing, 2015.
- [25] S. B. Kotsiantis, D. Kanellopoulos and P. Pintelas, "Data preprocessing for supervised learning," *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006.
- [26] M. L. Giger and K. Suzuki, "Computer-aided diagnosis," in *Biomedical Information Technology*, New York, NY, USA: Academic Press, pp. 359–374, 2008.
- [27] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4–5, pp. 198–211, 2008.
- [28] A. Cahan and J. J. Cimino, "A learning health care system using computer-aided diagnosis," *Journal of Medical Internet Research*, vol. 19, no. 3, pp. e54, 2017.
- [29] I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective," *Artificial Intelligence and Medicine*, vol. 23, no. 1, pp. 89–109, 2001.
- [30] M. Janda and H. P. Soyer, "Can clinical decision making be enhanced by artificial intelligence?," *British Journal of Dermatology*, vol. 180, no. 2, pp. 247–248, 2019.
- [31] C. W. Hanson and B. E. Marshall, "Artificial intelligence applications in the intensive care unit," *Critical Care Medicine*, vol. 29, no. 2, pp. 427–435, 2001.
- [32] P. Sajda, "Machine learning for detection and diagnosis of disease," *Annual Review of Biomedical Engineering*, vol. 8, no. 1, pp. 537–565, 2006.
- [33] A. Lal, Y. Pinevich, O. Gajic, V. Herasevich and B. Pickering, "Artificial intelligence and computer simulation models in critical illness," *World Journal of Critical Care Medicine*, vol. 9, no. 2, pp. 13–19, 2020.
- [34] M. Komorowski, "Artificial intelligence in intensive care: Are we there yet?," *Intensive Care Medicine*, vol. 45, no. 9, pp. 1298–1300, 2019.
- [35] P. Mathur and M. L. Burns, "Artificial intelligence in critical care," *International Anesthesiology Clinics*, vol. 57, no. 2, pp. 89–102, 2019.
- [36] L. Shinnars, C. Aggar, S. Grace and S. Smith, "Exploring healthcare professionals' understanding and experiences of artificial intelligence technology use in the delivery of healthcare: An integrative review," *Health Informatics Journal*, vol. 26, no. 2, pp. 1225–1236, 2020.

- [37] D. Denyer and D. Tranfield, "Producing a systematic review," in *The Sage Handbook of Organizational Research Methods*, D. A. Buchanan, A. Bryman, Thousand Oaks, USA: Sage, pp. 671–689, 2009.
- [38] K. S. Khan, R. Kunz, J. Kleijnen and G. Antes, "Five steps to conducting a systematic review," *Journal of the Royal Society of Medicine*, vol. 96, no. 3, pp. 118–121, 2003.
- [39] S. Gopalakrishnan and P. Ganeshkumar, "Systematic reviews and meta-analysis: Understanding the best evidence in primary healthcare," *Journal of Family Medicine and Primary Care*, vol. 2, no. 1, pp. 9–14, 2013.
- [40] S. L. Chong, N. Liu, S. Barbier and M. E. H. Ong, "Predictive modeling in pediatric traumatic brain injury using machine learning data analysis, statistics and modelling," *BMC Medical Research Methodology*, vol. 15, no. 1, pp. 1–9, 2015.
- [41] I. R. Abubakar and S. O. Olatunji, "Computational intelligence-based model for diarrhea prediction using demographic and health survey data," *Soft Computing*, vol. 24, no. 7, pp. 5357–5366, 2020.
- [42] R. Asaoka, H. Murata, A. Iwase and M. Araie, "Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier," *Ophthalmology*, vol. 123, no. 9, pp. 1974–1980, 2016.
- [43] Y. Raita, T. Goto, M. K. Faridi, D. F. M. Brown, A. C. Camargo *et al.*, "Emergency department triage prediction of clinical outcomes using machine learning models," *Critical Care*, vol. 23, no. 1, pp. 1–13, 2019.
- [44] B. Wang, Y. Bai, Z. Yao, J. Li, W. Dong *et al.*, "A multi-task neural network architecture for renal dysfunction prediction in heart failure patients with electronic health records," *IEEE Access*, vol. 7, pp. 178392–178400, 2019.
- [45] A. Elsayad and M. Fakhr, "Diagnosis of cardiovascular diseases with bayesian classifiers," *Journal of Computer Science*, vol. 11, no. 2, pp. 274–282, 2015.
- [46] K. Maghooli, M. Langarizadeh, L. Shahmoradi, M. Habibikoolaei, M. Jebraeily *et al.*, "Differential diagnosis of erythmato-squamous diseases using classification and regression tree," *Acta Informatica Medica*, vol. 24, no. 5, pp. 338–342, 2016.
- [47] A. J. Aljaaf, D. Al-Jumeily, H. M. Haglan, M. Alloghani, T. Baker *et al.*, "Early prediction of chronic kidney disease using machine learning supported by predictive analytics," in *IEEE Congress on Evolutionary Computation*, Rio de Janeiro, Brazil, pp. 1–9, 2018.
- [48] A. A. Nahid, N. Sikder, A. K. Bairagi, M. A. Razzaque, M. Masud *et al.*, "A novel method to identify pneumonia through analyzing chest radiographs employing a multichannel convolutional neural network," *Sensors*, vol. 20, no. 12, pp. 1–18, 2020.
- [49] O. Stephen, M. Sain, U. J. Maduh and D. U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *Journal of Healthcare Engineering*, vol. 2019, pp. 1–7, 2019.
- [50] Y. Horie, T. Yoshio, K. Aoyama, S. Yoshimizu, Y. Horiuchi *et al.*, "Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks," *Gastrointestinal Endoscopy*, vol. 89, no. 1, pp. 25–32, 2019.
- [51] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informatics*, vol. 5, no. 2, pp. 1–14, 2018.
- [52] X. Kong, S. Gong, L. Su, N. Howard and Y. Kong, "Automatic detection of acromegaly from facial photographs using machine learning methods," *EBioMedicine*, vol. 27, pp. 94–102, 2018.
- [53] R. Nijhawan, R. Verma, Ayushi, S. Bhushan and R. Dua, "An integrated deep learning framework approach for nail disease identification," in *13th Int. Conf. on Signal-Image Technology & Internet-Based Systems*, Jaipur, India, pp. 197–202, 2018.
- [54] M. A. Elaziz, K. M. Hosny, A. Salah, M. M. Darwish, S. Lu *et al.*, "New machine learning method for image based diagnosis of COVID-19," *PLoS One*, vol. 15, no. 6, pp. 1–18, 2020.
- [55] J. Krois, T. Ekart, L. Meinhold, T. Golla, B. Kharbot *et al.*, "Deep learning for the radiographic detection of periodontal bone loss," *Scientific Reports*, vol. 9, no. 1, pp. 1–6, 2019.
- [56] S. Murakami, K. Hatano, J. Tan, H. Kim and T. Aoki, "Automatic identification of bone erosions in rheumatoid arthritis from hand radiographs based on deep convolutional neural network," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 10921–10937, 2018.

- [57] I. Ionita and L. Ionita, "Prediction of thyroid disease using data mining techniques," *Broad Research in Artificial Intelligence and Neuroscience*, vol. 7, pp. 115–124, 2016.
- [58] M. P. Than, J. W. Pickering, Y. Sandoval, A. S. V. Shah, A. Tsanas *et al.*, "Machine learning to predict the likelihood of acute myocardial infarction," *Circulation*, vol. 140, no. 11, pp. 899–909, 2019.
- [59] R. A. Taylor, C. L. Moore, K. H. Cheung and C. Brandt, "Predicting urinary tract infections in the emergency department with machine learning," *PLoS One*, vol. 13, no. 3, pp. 1–15, 2018.
- [60] X. Tian, Y. Chong, Y. Huang, P. Guo, M. Li *et al.*, "Using machine learning algorithms to predict hepatitis B surface antigen seroclearance," *Computational and Mathematical Methods in Medicine*, vol. 2019, pp. 1–7, 2019.
- [61] L. Yan, H. T. Zhang, J. Goncalves, Y. Xiao, M. Wang *et al.*, "An interpretable mortality prediction model for Covid-19 patients," *Nature Machine Intelligence*, vol. 2, no. 5, pp. 283–288, 2020.
- [62] M. D. Ganggayah, N. A. Taib, Y. C. Har, P. Lio and S. K. Dhillon, "Predicting factors for survival of breast cancer patients using machine learning techniques," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1–17, 2019.
- [63] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju *et al.*, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in Genetics*, vol. 9, pp. 1–10, 2018.
- [64] P. Samant and R. Agarwal, "Machine learning techniques for medical diagnosis of diabetes using iris images," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 121–128, 2018.
- [65] C. C. Wu, M. M. Islam, T. Nasrin, P. A. Nguyen, H. C. Yang *et al.*, "Prediction of fatty liver disease using machine learning algorithms," *Computer Methods and Programs in Biomedicine*, vol. 170, pp. 23–29, 2019.
- [66] A. Leha, K. Hellenkamp, B. Unsold, S. M. Blake, A. M. Shah *et al.*, "A machine learning approach for the prediction of pulmonary hypertension," *PLoS One*, vol. 14, no. 10, pp. 1–16, 2019.
- [67] V. Thangaraj and V. Natarajan, "Glaucoma diagnosis using support vector machine," in *Proc. of 2017 Int. Conf. on Intelligent Computing and Control Systems*, Madurai, pp. 394–399, 2017.
- [68] A. Singh, M. K. Dutta, R. Jennane and E. Lespessailles, "Classification of the trabecular bone structure of osteoporotic patients using machine vision," *Computers in Biology and Medicine*, vol. 91, pp. 148–158, 2017.
- [69] M. Ahmad, V. Tundjungsari, D. Widiarti, P. Amalia and U. A. Rachmawati, "Diagnostic decision support system of chronic kidney disease using support vector machine," in *Proc. of 2nd Int. Conf. on Informatics and Computing*, Jayapura, Indonesia, pp. 1–4, 2018.
- [70] C. Sady and A. L. P. Ribeiro, "Symbolic features and classification via support vector machine for predicting death in patients with chagas disease," *Computers in Biology and Medicine*, vol. 70, pp. 220–227, 2016.
- [71] K. C. Hsu, C. H. Lin, K. R. Johnson, C. H. Liu, Y. T. Chan *et al.*, "Autodetect extracranial and intracranial artery stenosis by machine learning using ultrasound," *Computers in Biology and Medicine*, vol. 116, 103569, 2020.
- [72] E. Marquez and V. Barron, "Artificial intelligence system to support the clinical decision for influenza," in *IEEE Int. Autumn Meeting on Power, Electronics and Computing (ROPEC)*, Ixtapa, Mexico, pp. 1–5, 2019.
- [73] N. Hameed, A. Shabut and M. A. Hossain, "A computer-aided diagnosis system for classifying prominent skin lesions using machine learning," in *10th Computer Science and Electronic Engineering Conference*, Colchester, UK, pp. 186–191, 2019.
- [74] N. Louridi, M. Amar and B. El Ouahidi, "Identification of cardiovascular diseases using machine learning," in *7th Mediterranean Congress Telecommunications*, Fes, Morocco, pp. 1–6, 2019.
- [75] L. Ali, A. Niamat, J. A. Khan, N. A. Golilarz, X. Xingzhong *et al.*, "An optimized stacked support vector machines based expert system for the effective prediction of heart failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019.

- [76] F. Karimi-Alavijeh, S. Jalili and M. Sadeghi, "Predicting metabolic syndrome using decision tree and support vector machine methods," *ARYA Atheroscler*, vol. 12, no. 3, pp. 146–152, 2016.
- [77] M. Bhattacharya, D. Y. Lu, S. M. Kudchadkar, G. V. Greenland, P. Lingamaneni *et al.*, "Identifying ventricular arrhythmias and their predictors by applying machine learning methods to electronic health records in patients with hypertrophic cardiomyopathy (HCM-VAr-risk model)," *American Journal of Cardiology*, vol. 123, no. 10, pp. 1681–1689, 2019.
- [78] S. Mohan, C. Thirumalai and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.
- [79] L. Boldú, A. Merino, S. Alférez, A. Molina, A. Acevedo *et al.*, "Automatic recognition of different types of acute leukemia in peripheral blood by image analysis," *Journal Clinical Pathology*, vol. 72, no. 11, pp. 755–761, 2019.
- [80] S. Gambhir, S. K. Malik and Y. Kumar, "PSO-ANN based diagnostic model for the early detection of dengue disease," *New Horizons in Translational Medicine*, vol. 4, no. 1–4, pp. 1–8, 2017.